

act_report

July 2, 2018

0.1 Analyzing, and Visualizing WeRateDogs Dataset after Wrangling process

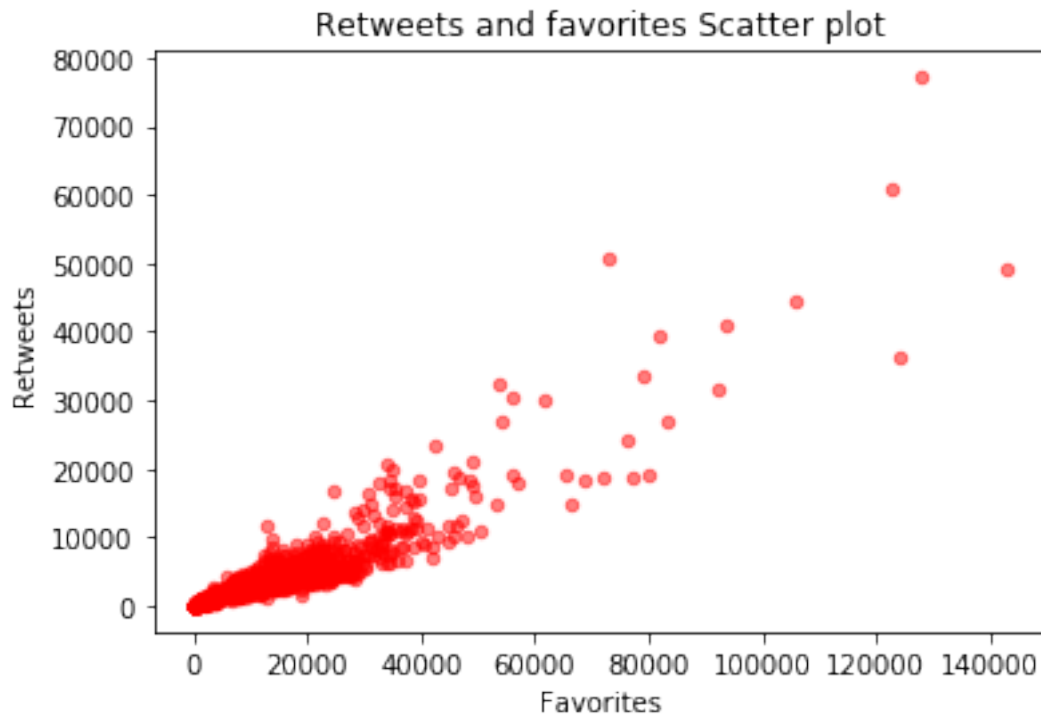
0.2 Introduction

You've no doubt heard of We Rate Dogs by now. Either you follow the account on Twitter, or you've read about the "They're good dogs Brent" story, which took the Internet by storm in 2016. In case you didn't WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. It was started in 2015 by college student Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter. Over the course of two years, creator and curator Matt Nelson has built an empire out of being the internet's go-to dog-rater. Nelson bestows rarely high ratings — on a 1-to-10 scale, almost all of his ratings are higher than 10 — on user-submitted pups, along with witty, ironic captions that are cute, hilarious, and relatable all at once. WeRateDogs started from humble beginnings — he claims the idea was hatched at an Applebee's in 2015 — and became an instant success. As of December 2017, the Twitter account has nearly 4.56 million followers. WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for us. Additional gathering, then assessing and cleaning was required to present these analyses and visualizations.

0.3 Favourites vs Retweets

```
In [14]: df_master.plot(kind = 'scatter', x = 'favorites', y = 'retweets', alpha = 0.5, color =
plt.xlabel('Favorites')
plt.ylabel('Retweets')
plt.title('Retweets and favorites Scatter plot')
```

```
Out[14]: Text(0.5,1,'Retweets and favorites Scatter plot')
```

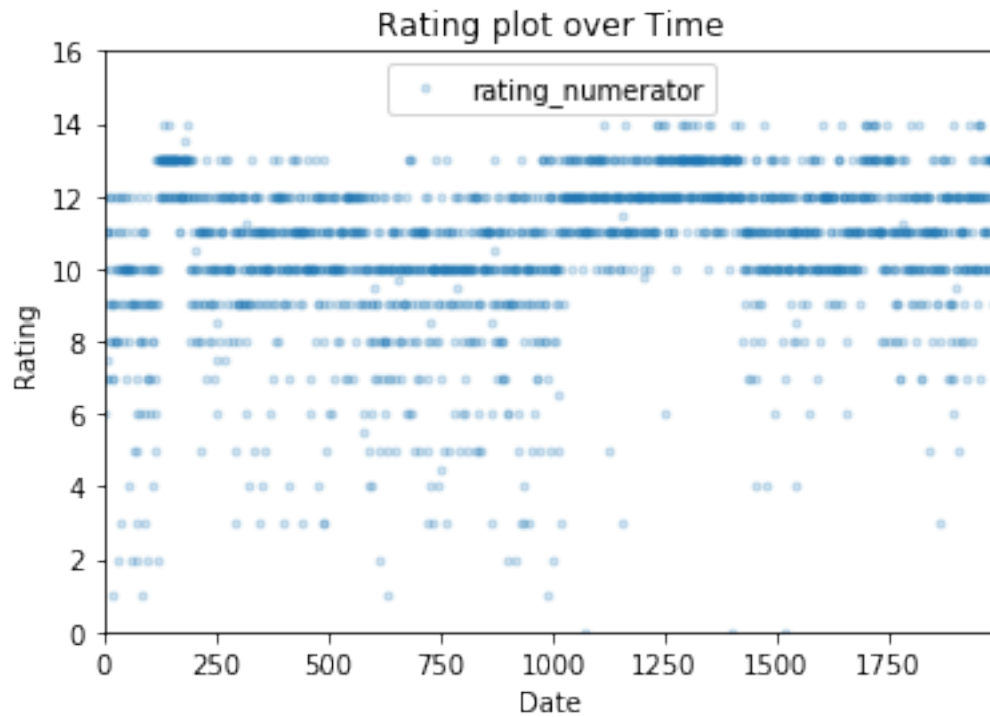


As the correlation map shows if the count of retweet is high for a specific tweet the count of favorites goes high as well. - The most retweeted post was a doggo realizing that he can stand in pool with 78,809 retweets - The most favorited post was a dog marching in the 2017 Women's March with 131,903 favorites

0.4 Rating Plot over time

```
In [15]: df_master.plot(y='rating_numerator', ylim=[0,16], style = '.', alpha = .2)
plt.title('Rating plot over Time')
plt.xlabel('Date')
plt.ylabel('Rating')
```

```
Out[15]: Text(0,0.5,'Rating')
```

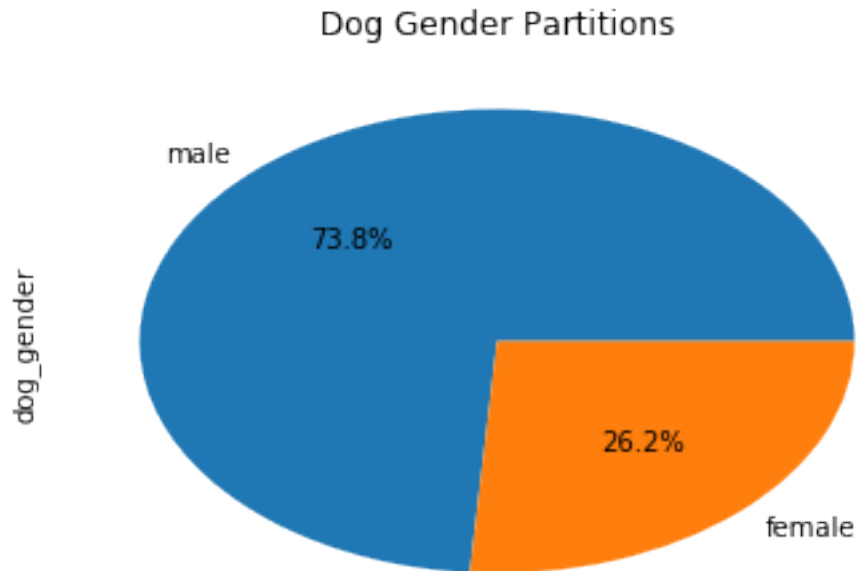


- More than 75% of the data has more than 12/10 as rating
- The page start with small rating than they adopt the system of rating numerator more than the denominator

0.5 Dog Gender Partitions

```
In [10]: df_master[df_master['dog_gender'].notnull()]['dog_gender'].value_counts().plot(kind = '
plt.title('Dog Gender Partitions')
```

```
Out[10]: Text(0.5,1,'Dog Gender Partitions')
```



- We have male dogs more than female dogs in our dataset, yet the female rating mean more than the male rating mean

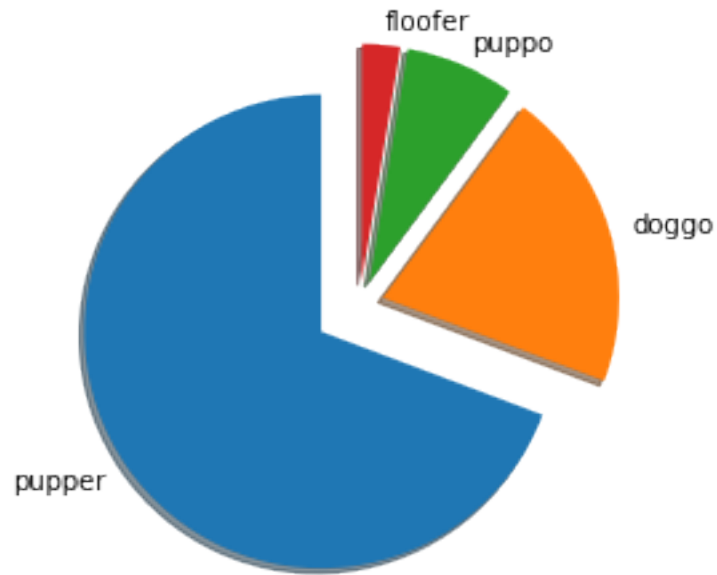
0.6 Dog Stages

The Four stages of Dog's life in our dataset are:

```
In [16]: dog_stage_count = list(df_master[df_master['dog_stage'] != 'None']['dog_stage'].value_counts())
dog_stages = df_master[df_master['dog_stage'] != 'None']['dog_stage'].value_counts().index
explode = (0.2, 0.1, 0.1, 0.1)

fig1, ax1 = plt.subplots()
ax1.pie(dog_stage_count, explode = explode, labels = dog_stages, shadow = True, startangle=90)
ax1.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
```

```
Out[16]: (-1.288268191449591,
1.2310305760057396,
-1.2401381220397572,
1.2110819987279693)
```



```
In [13]: df_master[['dog_gender', 'rating_numerator']][df_master.dog_gender.notnull()].groupby('dog_gender')
```

```
Out[13]:
```

	rating_numerator
dog_gender	
female	11.303097
male	10.652123

0.7 Conclusion

With the help of our clean dataset, we have managed to get the visualization on different parts of dataset. Therefore, we can say that our wrangling process is completed by looking into the visualization data and graphs.